

DATABASES

The Breast Cancer Information Core: Database Design, Structure, and Scope

Csilla Szabo,¹ Anthony Masiello,² Joseph F. Ryan,² The BIC Consortium,³ and Lawrence C. Brody^{2*}¹Unit of Genetic Epidemiology, International Agency for Research on Cancer, Lyon, France²Genetics and Molecular Biology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland³See Appendix

Communicated by Richard G.H. Cotton

The Breast Cancer Information Core (BIC) is an open access, on-line mutation database for breast cancer susceptibility genes. In addition to creating a catalogue of all mutations and polymorphisms in breast cancer susceptibility genes, a principle aim of the BIC is to facilitate the detection and characterization of these genes by providing technical support in the form of mutation detection protocols, primer sequences, and reagent access. Additional information at the site includes a literature review compiled from published studies, links to other internet-based, breast cancer information and research resources, and an interactive discussion forum which enables investigators to post or respond to questions and/or comments on a bulletin board. *Hum Mutat* 16:123–131, 2000. Published 2000 Wiley-Liss, Inc.[†]

KEY WORDS: BRCA1; BRCA2; breast cancer; mutation database; MDI

DATABASES:

BRCA1 – OMIM:113705; GDB:126611; GenBank:U14680; HGMD:BRCA1

BRCA2 – OMIM:600185; GDB:387848; GenBank:U43746; HGMD:BRCA2

INTRODUCTION

Breast cancer is the most common malignancy and the leading cause of cancer mortality among women worldwide [Parkin et al., 1999]. Approximately 5–10% of all breast cancer [Claus et al., 1991; Claus et al., 1996; Newman et al., 1988] and more than 30% of disease diagnosed under the age of 30 [Claus et al., 1991] is attributable to inheritance of a mutation in one or more highly penetrant autosomal dominant susceptibility genes. Identification and cloning of two such genes, BRCA1 (MIM# 113705) [Hall et al., 1990; Miki et al., 1994] and BRCA2 (MIM# 600185) [Tavtigian et al., 1996; Wooster et al., 1995; Wooster et al., 1994], rapidly led to the characterization of mutations in these genes among high-risk families, as well as breast and/or ovarian cancer patients or tumor series in populations worldwide [Castilla et al., 1994; Couch and Weber, 1996; Friedman et al., 1994; Neuhausen et al., 1996b; Phelan et al., 1996; Shattuck-Eidens et al., 1995; Simard et al., 1994; Thorlacius et al., 1996].

It was evident from the outset that widespread scientific and clinical interest in BRCA1 and BRCA2 would lead to an intensive effort to screen women at high risk of developing breast cancer for mutations in these genes. It is hoped that molecular information about cancer susceptibility genes will eventually be translated into clinical benefits. Building a knowledge base about cancer susceptibility genes is one of the steps towards realizing this goal. This seemed particularly likely to be a slow process

Received 10 November 1999; accepted revised manuscript 21 April 2000.

*Correspondence to: Lawrence C. Brody, Genetics and Molecular Biology Branch, National Human Genome Research Institute, 49 Convent Drive, Room 3A14, Bethesda, MD 20892-4442. E-mail: lbrody@helix.nih.gov

The BIC database is accessible through the World Wide Web site hosted by the National Human Genome Research Institute, USA: http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/Bic/

[†]This article is a US Government work and, as such, is in the public domain in the United States of America.

for the newly discovered BRCA1 and BRCA2 genes. Both encode large proteins with minimal sequence homology to known genes. Both appear to have a broad spectrum of putative pathologic mutations. Additionally, both have variant alleles found at high frequency in the population but not associated with a dramatic increase in cancer risk.

Coincident with the discovery of the BRCA1 and BRCA2 genes were changes in the way genetic information was published and accessed. As with DNA sequence information, the conventional scientific literature is no longer the most efficient venue for publishing mutation data. This is especially true of reports containing information obtained from one or a few families. Nevertheless, these "single mutation" reports can have value in aggregate, which is maximized by their distribution in a freely available format.

Hence, there was a need to make available to the research community information to aid in the identification of mutations and to provide a mechanism for collecting and distributing mutation data from the research community as a whole. A coordinated international collaborative effort was undertaken to compile information on the BRCA1 and BRCA2 genes. In addition to descriptions of the mutations found in the BRCA1 and BRCA2 genes, a principal objective of this consortium was to provide information on mutation detection techniques and gene specific primers in order to help those seeking to develop mutation-screening efforts. These data were made widely accessible through development of the Breast Cancer Information Core (BIC) with oversight provided by a curator (L.B.) and steering committee (see Appendix). [Friend et al., 1995].

Since its establishment in 1995 more than 1,700 members of the research community representing 48 different countries have registered to use the BIC database. After the expected initial high volume of registrants, new members have continued to register at a sustained rate of approximately five per week for several years (Fig. 1). This wide geographic representation and continued influx of new members suggests that the database is reaching an expanding audience.

As of April 2000, the site contains: 3,416 entries describing genetic variants in BRCA1 and 2,292 entries for BRCA2 (Fig. 2); a compilation of BRCA1 and BRCA2 intronic sequences and recommended PCR primers; a collection of eight detailed protocols describing diverse methods for detecting mutations in these genes. The scope, format, and organisation of the BIC database are

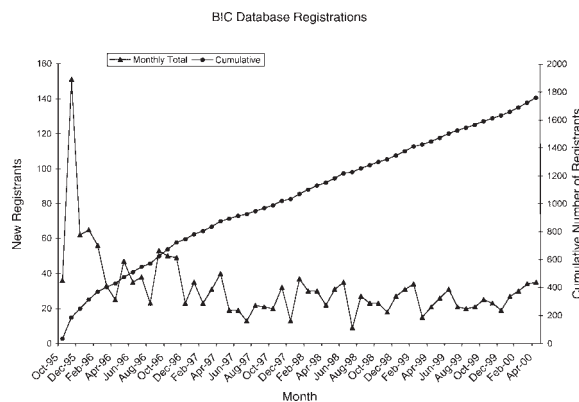


FIGURE 1. BIC member registrations. Number of individuals registering to access the BIC database per month (left axis) and total number registered (right axis) since the debut of the database on the World Wide Web.

described herein. Technical details about the scripts and software used in the database have not been included but are available upon request. An in depth description and analysis of the specific BRCA1 and BRCA2 mutation information and its inherent biases is under preparation.

DESCRIPTION/DATA STRUCTURE

Access

Registration for membership in the BIC is requested to ensure that individuals using the database agree to a set of guidelines covering participation, data entry, confidentiality, as well as appropriate data use and acknowledgement. Data are accessible to members through a password provided immediately upon registration. Database guidelines and registration requirements are not found in the majority of other locus specific da-

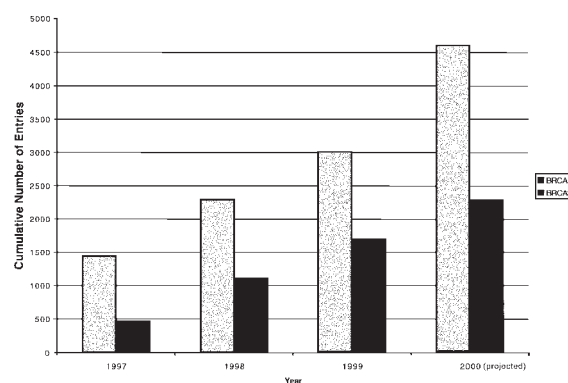


FIGURE 2. Cumulative number of mutation entries for BRCA1 and BRCA2 in the BIC database. The data is current as of April 2000. Year 2000 data is projected assuming data will be deposited at the rate it was received from January through April of this year. Entries received by the curator but not yet processed are included.

tabases. The BIC has maintained this requirement as a mechanism to protect the contributors of unpublished information and to remind users that the data is designed primarily as a research resource and is not appropriate for clinical decision making. All registrants are given access to the database.

Data Source

The BIC database is a compilation of data derived from both published literature and direct on-line entries contributed by researchers throughout the world. Built-in redundancies in data entry information (e.g., descriptors of the mutation at both the nucleotide and amino acid level) facilitate data validation by the database curator prior to posting on the web site. At present, data is checked manually by the curator. Implementation of an automated data validity checking tool, such as DNA Mutation Checker (<http://www2.ebi.ac.uk/cgi-bin/mutations/check.cgi>; H. Lehtväslaiho), is being explored.

To date, the database includes BRCA1 and BRCA2 mutations identified in germline DNA of breast and/or ovarian cancer patients from either high-risk families or patient series. Germline and somatic mutations described in tumors of the breast, ovary, and other tissues are also included. The allelic frequency of polymorphic variants is given for normal control chromosomes when known. However, since only information from probands and tumors in which mutations have been identified is recorded in the BIC, it is not possible to determine from these data what proportion of high-risk families, systematic series of patients, or tumor samples are attributable to mutations in either gene.

Data Structure

The overall layout of the BIC web site is shown in Figure 3. Separate mutation databases exist within the BIC site for the BRCA1 and BRCA2 genes. Each gene is depicted graphically, with hyperlinks between features enabling the user to shuttle between the gene image and exon-specific mutation records. These are structured as flat files containing 24 fields for each entry (See Table 1). A condensed database format is also available in which the mutations reported for each exon are summarised according to the type of genetic alteration (e.g., frameshift, nonsense, missense, splice). An overall summary sheet describes the total number of alterations reported in each exon and the number of these variants that are distinct, as well as the number represented by a single entry only.

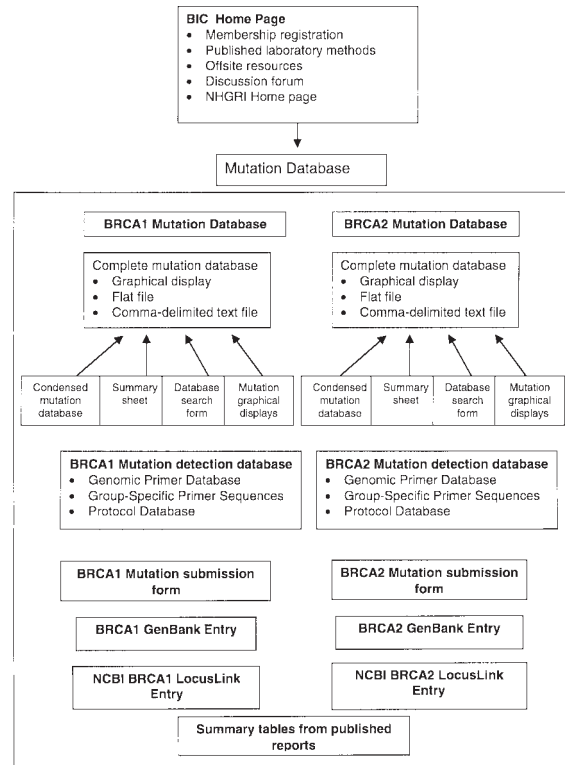


FIGURE 3. The Breast Information Core (BIC) Website Features. Overview and map of the BIC website. Details of the links contained in the Mutation Database section are shown.

It is possible to use hyperlinks to jump between the summary tables and the comprehensive data tables. The entire database for each gene is available for download as a comma-delimited text file. In addition, using the built-in dynamic search engine, it is possible to search the databases for specific information contained in 20 of the data fields.

A series of data visualization tools were recently developed. 1) A substitution matrix indicates at what frequency single nucleotide substitutions occur between bases. 2) Insertion and deletion graphs reveal the number of times a base is reported as an insertion or deletion. 3) A frequency graph shows how many mutations were reported at each cDNA base. Each type of mutation (frameshift, missense, nonsense, polymorphism, splice, and unclassified) is given its own color. The default view shows all types of mutations, but each classification can be viewed in isolation. Clicking on the graph zooms into the specific area that was clicked. 4) The distinct mutation graph plots a single bar wherever a type of mutation was reported. Like the frequency graph, classifications can be isolated and it is possible to zoom in on areas. 5) The stop codon and stop cDNA graphs reveal where stop-

TABLE 1. BIC Mutation Database Fields

| | |
|-------------------------------------|--|
| Accession Number | Unique identifier generated at the time mutation is entered into database |
| Exon | BRCA1 or BRCA2 exon in which mutation has been identified |
| cDNA nucleotide | Nucleotide # in the transcript at which mutation occurs; Reference sequences: BRCA1 GenBank U14680; BRCA2 GenBank U43746 |
| gDNA nucleotide | Nucleotide # in genomic DNA at which mutation occurs; Reference sequences: BRCA1 GenBank L78833 |
| Codon | Triplet codon # (ATG is +1) in which mutation occurs |
| Base change | Description of nucleotide difference compared to reference sequence |
| Amino acid change | Description of resulting change in the encoded amino acid sequence |
| Mutation designation | Designation of described mutation according to nomenclature guidelines |
| Mutation type | Frameshift, nonsense, missense, splice, in frame deletion |
| Mutation effect | Frameshift, nonsense, missense, splice, unclassified variant, polymorphism |
| Depositor | Contributor of mutation report |
| Patient sample source | Hospital or research institute from which patient sample originated |
| Patient ID number | Patient identifier designation in publications, or used by hospital or research institute |
| Number of times reported | Number of family members carrying mutation |
| Germline or Somatic mutation | |
| Detection method used | SSCP, direct sequencing, DGGE etc. |
| Proband tumor type | Breast, ovarian, other |
| Number of chromosomes screened | Number of normal control chromosomes screened |
| Frequency of polymorphism (A/C/G/T) | Information on the frequency of variants in normal control chromosomes, if known |
| Literature reference | Literature citation in which mutation was first reported |
| Contact person | Email address of individual to whom inquiries should be addressed |
| Notes | Additional information describing the mutation, family history, patient series, age of onset, etc. |
| Creation date | Date on which mutation is entered into the public database |
| Ethnicity | Information on patient ethnicity, if known |
| Nationality | Information on patient nationality, if known |

causing mutations occur. These graphs also have zoom capability.

Although each piece of software is custom written, the modular design of the database allows the current structure to be "cloned" and expanded to new loci as appropriate. For example, the original database only covered BRCA1. BRCA2 information was added soon after the gene was cloned by duplicating and modifying the appropriate sections from the BRCA1 database.

Software

The initial planning stages of the BIC database coincided with the widespread development of information display on the World Wide Web. Web-based display of data circa 1995 seems quite primitive when judged by current standards. The BIC began as static HTML pages that had been manually converted from a spreadsheet. Data was collected either by e-mail or by postal submissions. When they became available, HTML forms (used to submit information directly from a web page) were added to assist in data collection. This allowed depositors to submit mutations using a standard template and helped make the collected data more consistent. The data from the submission form was then e-mailed to the curator who examined submissions and entered the data into the

spreadsheet. The new entries in the spreadsheet were subsequently manually appended to the static HTML pages.

As the technology matured and the volume of data grew, the data was moved from static HTML files to a Sybase database server. This migration helped centralize the data and simplified the updating process. Accession numbers and entry date could now be added to each entry automatically. In addition, it became possible to create applications utilizing the database's built-in data mining tools.

Software Tools

Using SQL (Structured Query Language) and PERL (Practical Extraction and Report Language), a suite of web accessible tools have been developed to aid analysis of the BIC data. These tools are used to generate a summary page presenting a broad overview of the data. Custom scripts were written to extract the total number of mutations reported for each exon as well as how many of those entries were "distinct" (see below) and how many were reported only once in the entire database. A search engine is provided for searching based on most database fields. A clickable image map allows for quick access to the data by exon. Recently, series of database queries were written to extract information from the database and represent it in graphical form. This

program generates a large series (>10,000) of static HTML images hyperlinked to each other to allow the user to effectively zoom in from a global to small scale view of the data (Fig. 4). The display of static images (essentially cached queries) was chosen over having the graphs generated "on the fly" to minimize the time a user spends waiting for the information to be returned. While this approach requires a considerable amount of storage space, it has the added advantage of reducing processor time.

Future tools to be added will provide links between the currently available views of the data. For example, while it is now possible to move from a summary graph or table to an individual entry, one cannot jump between graphical and tabular displays. Work is currently in progress to provide hyperlinks from the graphs to the full records in the database. Beyond this, we are now planning to increase the "resolution" of the data displayed to the nucleotide level. Scripts are being written to generate a new theoretical sequence file for each mutation. These would then be displayed on the site aligned to the reference sequence. Initially, cDNA alignments will

be displayed; alignments to full genomic sequences will follow. The capacity to display actual sequence information will be useful for resolving mutation nomenclature issues (see below). This feature will be incorporated into the mutation entry form to facilitate automated data checking through alignment of a short nucleotide sequence flanking the mutation (furnished by the depositor) with the reference sequence. Nonetheless, users will be cautioned that the sequence display is purely theoretical and not a representation of the actual DNA sequence of the individual with the mutation.

PROBLEMS

Nomenclature Issues

The BIC database has adopted most of the nomenclature system recommendations of the HUGO Nomenclature Working Group [Antonarakis, 1998] for designation of mutations in the BRCA1 and BRCA2 genes as derived from the GenBank reference sequences U14680 and U43746, respectively. One significant departure from the HUGO Working Group recommendations



FIGURE 4. Graphical display of mutation data. A schematic of the cDNA with the location of each mutation indicated is generated directly from the mutation data. Each mutation type is assigned a specific color. Users may click a region of the graph to zoom in on an area of interest. Each new window represents a 10-fold magnification. Clicking on a specific mutation in the key restricts the display to that mutation type only. The vertical purple arrow marks the same nucleotide in each window.

is that the cDNA number system used by the BIC designates the first nucleotide of the GenBank entry as nucleotide #1. For BRCA1, this entry includes 119 bases of 5' untranslated sequence with protein translation starting at nucleotide #120. The reference BRCA2 sequence contains 5' untranslated region of 228 nucleotides. The untranslated segment must be taken into account when converting from cDNA positions to codon number. The HUGO system recommends using the "A" of the ATG start codon as the number 1 position. While the HUGO system can accommodate multiple GenBank entries with different 5' untranslated regions, all of the BIC data is based on a single reference sequence for each gene. These reference sequences are stable entities, should they be edited in the future, the National Center for Biotechnology Information, the curators of GenBank, plans to retain the original sequences in the database. In the future a field will be added to the current database containing the equivalent HUGO Nomenclature Working Group approved designation for each mutation. This will allow easy exchange between the BIC database and the proposed HUGO Core Database.

The single greatest source of discrepancy among designations for entries describing (apparently) identical mutations is the ambiguous labelling of the nucleotide position of insertions or deletions consisting of unit(s) within short tandem repeats (STR: mono-, di-, tri-, etc. nucleotide tracts). Hence, the designation for the BRCA1 185delAG mutation, which deletes one of two tandem AG di-nucleotide repeats, describes the same mutation (at the nucleotide level) as 187delAG. While it is not possible to determine which of the AG repeats was lost during the mutational event, the consequences are the same (frameshift at amino acid 23; truncation at amino acid 39). When possible, the BIC adheres to the HUGO recommendations in retaining as much of the "normal" sequence as possible before the deleted nucleotide.

To create the software tools required for generating the data summary sheets and graphical displays, mutation reports that contain identical designations in several data fields are considered to describe the same mutation, enabling a count of the number of times the same mutation has been reported to the database. A new term, *distinct mutation* was coined for classification purposes. Two entries are distinct (from each other) if they differ in key database fields. This avoids the problem of having to distinguish between true recurrent mutations (identical by state) and founder mutations (identical by descent).

In order for a count of distinct mutations in each

gene to be accurate, it is essential that mutations identical at the nucleotide level be entered with a consistent nomenclature. However, several of the more frequent mutations were identified prior to development of the current nomenclature guidelines. Hence, the ambiguity in defining which repeat of an STR was deleted, or the point of insertion of a repeat within an STR, is reflected in the naming of several of the more frequently reported mutations (e.g., the BRCA1 mutations 185delAG and 5382insC would be designated 187delAG and 5385insC under current nomenclature guidelines). Renaming such mutations at this late date may inadvertently create confusion in the literature with respect to these mutations. Therefore, the original names of these "common" mutations are retained in the BIC database in the 'Mutation Designation' data field. All other mutations in repeat regions are changed to conform to the standard.

Recent identification of mutations in both BRCA1 [Montagna et al., 1999; Petrij-Bosch et al., 1997; Puget et al., 1999; Puget et al., 1997; Swensen et al., 1997] and BRCA2 [Nordling et al., 1998] arising through large genomic rearrangements has underscored the necessity of designating mutation nomenclature not only with respect to the reference cDNA sequence, but also to the reference genomic DNA sequence(s) (e.g., BRCA1 GenBank L78833, [Smith et al., 1996]). This will enable descriptions of such rearrangements to be specified with accuracy to the nucleotide level. Updating of the mutation tables to include genomic DNA mutation designations is under way.

Comprehensiveness

Although the BIC curator and Steering Committee members make a good faith effort to incorporate all published mutations, the BIC does not have a formal journal scanning system. The rapid proliferation of such reports essentially leads to the database lagging somewhat behind the literature. To circumvent this delay in data entry, we suggest that authors submit mutations to the database concurrently with manuscript acceptance and/or that editors of journals alert the BIC curator to newly accepted manuscripts detailing mutations in the BRCA1 and BRCA2 genes.

In addition to mutation descriptions, the BIC provides data fields for associated information. Contributors are asked to provide Proband Tumor Type, Number of Chromosomes Screened, and Ethnicity or Nationality, as well as information for the frequency of specific polymorphic variants and information about mutation detection methodology. These data

fields are frequently incomplete in mutation entry submissions. The utility of the database would be greatly enhanced if contributors could endeavour to provide as complete information as possible.

Redundancy

As information for the database is culled from several different sources, it is important to be able to identify instances when the same alteration in the same individual/tumor sample has been entered into the database multiple times (i.e., literature report and direct submission). Thus, the importance of linking each described mutation with a unique individual/sample identifier by investigators, both in the literature and in submissions to the BIC database, must be emphasised.

Timeliness

As new features become available, the web site structure and the appearance is updated. Links to related information sites are checked approximately twice a year. New mutation data is delivered to the curator as the depositor enters it. Roughly every other month, accumulated entries are compiled, checked for discrepancies and transferred in bulk by the curator to the database. Therefore, depending on when in the cycle they are entered, some entries may appear in the database within days of receipt by the curator while more than a month may pass before other entries appear. An e-mail notice is sent to the depositor when his/her data has been posted.

The "discussion board" section of the database was conceived as an alternative to a standard Internet discussion group. Members post queries or findings in one of several broad categories. In theory, interested BIC members would periodically scan the postings, respond to the queries or share their own experiences. In practice the volume of information exchange occurring on the discussion board has yet to reach a self-sustaining critical mass. Queries would be posted and responded to many weeks later, if at all. The failure of this section to serve BIC members may reflect lack of interest or awareness on the part of database users. Alternatively, members may assume that "someone else will answer that question." A possible solution to this problem would be to appoint a moderator who would monitor the messages posted and try to match those seeking information with appropriate sources.

Stability

Since its debut, the BIC has had a single "curator" and a part-time staff of web and database programmers. As is true for the majority of the

locus-specific databases, the BIC content is created and curated by interested and dedicated individuals in their "spare time." This has led to some concern that locus-specific databases would appear and disappear from the web without notice. The Intramural research program at NHGRI supports these efforts by providing programmer time, computer hardware, data storage space, and web access. Some of the BIC information pages actually reside on servers elsewhere (BCLC Information, University of Leiden, P. Devilee). The BIC Steering Committee consists of a rotating pool of 10–12 investigators familiar with the field of breast cancer genetics who meet monthly via teleconference to discuss issues of relevance to maintaining and updating the database. This combination of a large steering committee, a degree of decentralisation, and commitment of institutional support should provide long term stability to the BIC. Special arrangements would have to be made if there was a change of curators or the current curator was to change institutions. In contrast, the Central Databases (e.g., GenBank, EBI, HGMD) are less vulnerable to this type of disruption.

FUTURE PROSPECTS

Mutation Databases of Additional Breast Cancer Susceptibility Genes

Other genes known to contribute to elevated breast cancer risk include p53 [Phillips, 1999], ATM [Meyn, 1999], and PTEN [Eng, 1998], and associations with elevated risk and rare variants of the HRAS locus VNTR have been frequently reported [Conway et al., 1995]. Among these, independent databases have been developed for both the p53 gene and the ATM gene [Soussi et al., 2000] (links to these and other locus-specific databases are maintained by the Human Gene Mutation Database, Cardiff: www.uwcm.ac.uk/uwcm/mg/hgmd0.html) [Krawczak et al., 2000]. Links to these particular locus-specific databases are included in the BIC. In addition, a mutation database for the PTEN gene, which has been shown to contribute to an elevated risk of breast cancer in the context of Cowden's syndrome [Eng, 1998], is under construction and will be added to the BIC web site in the near future.

Haplotype Database

A number of mutations have been found to occur in multiple, unrelated individuals [Neuhausen et al., 1996a; Oddoux et al., 1996; Petrij-Bosch et al., 1997; Roa et al., 1996; Simard et al., 1994; Struewing et al., 1995; Thorlacius et al., 1996]. In

most instances, this frequent occurrence is attributable to a single ancestral mutation exhibiting a strong founder effect in specific populations [Neuhausen et al., 1998; Neuhausen et al., 1996b; reviewed in Szabo and King [1997]. A database containing haplotypes associated with founder mutations is being constructed. This database will be structured such that existing microsatellite marker data and newer single nucleotide polymorphism (SNP) based data can be incorporated.

SUMMARY

The BIC database currently provides the research community with a central repository of mutations and polymorphisms characterized in the BRCA1 and BRCA2 breast cancer susceptibility genes. Since its establishment in 1995, the number of entries for each gene has increased at a constant rate of approximately 500–600/year (Fig. 2). It is notable that despite the large volume of mutations already described for these genes, novel variants continue to be identified. The proportion of entries derived from direct submission of unpublished work continues to increase suggesting that electronic locus-specific databases, such as the BIC, serve a critical function in recording and providing access to these data while publication of mutation reports declines. As these data accrue, it is hoped that the compendium of mutations will provide a rich resource from which to derive hypotheses for future research avenues probing gene function, or differences in risks, clinical prognosis and/or responsiveness to different therapeutic interventions that might be associated with specific mutations.

ACKNOWLEDGMENTS

The authors would like to thank L. Cozzens, R. Watson, J. Ramacciato, E. Tachibana, M. McDonald, and G. Birznies for programming assistance, S. Danoff for critically reading the manuscript, and numerous BIC members who have suggested improvements in the database.

REFERENCES

- Antonarakis SE. 1998. Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum Mutat* 11:1–3.
- Castilla LH, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE, Boyd J, Lubin MB, Deshano ML, Brody LC, Collins FS, Weber BL. 1994. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat Genet* 8:387–391.
- Claus EB, Risch N, Thompson WD. 1991. Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am J Hum Genet* 48:232–242.
- Claus EB, Schildkraut JM, Thompson WD, Risch NJ. 1996. The genetic attributable risk of breast and ovarian cancer. *Cancer* 77:2318–2324.
- Conway K, Edmiston S, Fried DB, Hulka BS, Garrett PA, Liu ET. 1995. Ha-ras rare alleles in breast cancer susceptibility. *Breast Cancer Res Treat* 35:97–104.
- Couch FJ, Weber BL. 1996. Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene. *Breast Cancer Information Core. Hum Mutat* 8:8–18.
- Eng C. 1998. Genetics of Cowden syndrome: through the looking glass of oncology. *Int J Oncol* 12:701–710.
- Friedman LS, Ostermeyer EA, Szabo CI, Dowd P, Lynch ED, Rowell SE, King MC. 1994. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nat Genet* 8:399–404.
- Friend S, Borresen AL, Brody L, Casey G, Devilee P, Gayther S, Goldgar D, Murphy P, Weber BL, Wiseman R. 1995. Breast cancer information on the web [letter]. *Nat Genet* 11:238–239.
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. 1990. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250:1684–1689.
- Krawczak M, Ball EV, Fenton I, Stenson PD, Aboysinghe S, Thomas N, Cooper DN. 2000. The human gene mutation database: a biomedical information and research resource. *Hum Mutat* 15:45–51.
- Meyn MS. 1999. Ataxia-telangiectasia, cancer and the pathobiology of the ATM gene. *Clin Genet* 55:289–304.
- Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian K, Liu Q, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S, Cannon-Albright L, Goldgar D, Wiseman R, Kamb A, Skolnick MH. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266:66–71.
- Montagna M, Santacatterina M, Torri A, Menin C, Zullato D, Chicco-Bianchi L, D'Andrea E. 1999. Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families. *Oncogene* 18:4160–4165.
- Neuhausen S, Gilewski T, Norton L, Tran T, McGuire P, Swensen J, Hampel H, Borgen P, Brown K, Skolnick M, Shattuck-Eidens D, Jhanwar S, Goldgar D, Offit K. 1996a. Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish women affected by breast cancer. *Nat Genet* 13:126–128.
- Neuhausen SL, Mazoyer S, Friedman L, Stratton M, Offit K, Caligo A, Tomlinson G, Cannon-Albright L, Bishop T, Kelsell D, Solomon E, Weber B, Couch F, Struwing J, Tonin P, Durocher F, Narod S, Skolnick MH, Lenoir G, Serova O, Ponder B, Stoppa-Lyonnet D, Easton D, King MC, Goldgar DE. 1996b. Haplotype and phenotype analysis of six recurrent BRCA1 mutations in 61 families: results of an international study. *Am J Hum Genet* 58:271–280.
- Neuhausen SL, Godwin AK, Gershoni-Baruch R, Schubert E, Garber J, Stoppa-Lyonnet D, Olah E, Csokay B, Serova O, Laloo F, Osorio A, Stratton M, Offit K, Boyd J, Caligo MA, Scott RJ, Schofield A, Teugels E, Schwab M, Cannon-Albright L, Bishop T, Easton D, Benitez J, King MC, Ponder BAJ, Weber B, Devilee P, Borg A, Narod S, Goldgar D. 1998. Haplotype and phenotype analysis of nine recurrent BRCA2 mutations in 111 families: results of an international study. *Am J Hum Genet* 62:1381–1388.
- Newman B, Austin MA, Lee M, King MC. 1988. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc Natl Acad Sci USA* 85:3044–3048.
- Nordling M, Karlsson P, Wahlstrom J, Engwall Y, Wallgren A, Martinsson T. 1998. A large deletion disrupts the exon 3 tran-

- scription activation domain of the BRCA2 gene in a breast/ovarian cancer family. *Cancer Res* 58:1372–1375.
- Oddoux C, Struewing JP, Clayton CM, Neuhausen S, Brody LC, Kaback M, Haas B, Norton L, Borgen P, Jhanwar S, Goldgar D, Ostrer H, Offit K. 1996. The carrier frequency of the BRCA2 6174delT mutation among Ashkenazi Jewish individuals is approximately 1%. *Nat Genet* 14:188–190.
- Parkin DM, Pisani P, Ferlay J. 1999. Estimates of the worldwide incidence of 25 major cancers in 1990. *Int J Cancer* 80:827–841.
- Petrij-Bosch A, Peelen T, van Vliet M, van Eijk R, Olmer R, Drusedau M, Hogervorst FB, Hageman S, Arts PJ, Ligtenberg MJ, Meijers-Heijboer H, Klijn JG, Vasen HF, Cornelisse CJ, van't Veer LJ, Bakker E, van Ommen GJ, Devilee P. 1997. BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat Genet* 17:341–345.
- Phelan CM, Lancaster JM, Tonin P, Gumbs C, Cochran C, Carter R, Ghadirian P, Perret C, Moslehi R, Dion F, Faucher MC, Dole K, Karimi S, Foulkes W, Lounis H, Warner E, Goss P, Anderson D, Larsson C, Narod SA, Futreal PA. 1996. Mutation analysis of the BRCA2 gene in 49 site-specific breast cancer families. *Nat Genet* 13:120–122.
- Phillips HA. 1999. The role of the p53 tumour suppressor gene in human breast cancer. *Clin Oncol* 11:148–155.
- Puget N, Torchard D, Serova-Sinilnikova OM, Lynch HT, Feunteun J, Lenoir GM, Mazoyer S. 1997. A 1-kb Alu-mediated germ-line deletion removing BRCA1 exon 17. *Cancer Res* 57:828–831.
- Puget N, Stoppa-Lyonnet D, Sinilnikova OM, Pages S, Lynch HT, Lenoir GM, Mazoyer S. 1999. Screening for germ-line rearrangements and regulatory mutations in BRCA1 led to the identification of four new deletions. *Cancer Res* 59:455–461.
- Roa BB, Boyd AA, Volcik K, Richards CS. 1996. Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. *Nat Genet* 14:185–187.
- Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod S, Couch F, Hoskins K, Weber B, Castilla L, Erdos M, Brody L, Friedman L, Ostermeyer E, Szabo C, King M-K, Jhanwar S, Offit K, Norton L, Gilewski T, Lubin M, Osborne M, Black D, Boyd M, Steel M, Ingles S, Haile R, Lindblom A, Olsson H, Borg A, Bishop DT, Solomon E, Radice P, Spatti G, Gayther S, Ponder B, Warren W, Stratton M, Liu Q, Fujimura F, Lewis C, Skolnick MH, Goldgar DE. 1995. A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene. Implications for presymptomatic testing and screening. *JAMA* 273:535–541.
- Simard J, Tonin P, Durocher F, Morgan K, Rommens J, Gingras S, Samson C, Leblanc JF, Belanger C, Dion F, et al. 1994. Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families. *Nat Genet* 8:392–398.
- Smith TM, Lee MK, Szabo CI, Jerome N, McEuen M, Taylor M, Hood L, King MC. 1996. Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. *Genome Res* 6:1029–1049.
- Soussi T, Dehouche K, Beroud C. 2000. The p53 mutation database. *Hum Mutat* 15:105–113.
- Struewing JP, Abeliovich D, Peretz T, Avishai N, Kaback MM, Collins FS, Brody LC. 1995. The carrier frequency of the BRCA1 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals. *Nat Genet* 11:198–200.
- Swensen J, Hoffman M, Skolnick MH, Neuhausen SL. 1997. Identification of a 14 kb deletion involving the promoter region of BRCA1 in a breast cancer family. *Hum Mol Genet* 6:1513–1517.
- Szabo CI, King MC. 1997. Population genetics of BRCA1 and BRCA2. *Am J Hum Genet* 60:1013–1020.
- Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S, Janecki T, Jiang P, Kehrer R, Leblanc JF, Mitchell JT, McArthur-Morrison J, Nguyen K, Peng Y, Samson C, Schroeder M, Snyder SC, Steele L, Stringfellow M, Stroup C, Swedlund B, Swensen J, Teng D, Thomas A, Tran T, Tranchant M, Weaver-Feldhaus J, Wong AKC, Shizuya H, Eyfjord JE, Cannon-Albright L, Labrie F, Skolnick MH, Weber B, Kamb A, Goldgar DE. 1996. The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat Genet* 12:333–337.
- Thorlacius S, Olafsdottir G, Tryggvadottir L, Neuhausen S, Jonasson JG, Tavtigian SV, Tulinius H, Ogmundsdottir HM, Eyfjord JE. 1996b. A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. *Nat Genet* 13:117–119.
- Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D, Fields P, Marshall G, Narod S, Lenoir GM, Lynch H, Feunteun J, Devilee P, Cornelisse CJ, Menko FH, Daly PA, Ormiston W, McManus R, Pye C, Lewis CM, Cannon-Albright LA, Peto J, Ponder BAJ, Skolnick MH, Easton DF, Goldgar DE, Stratton MR. 1994. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. *Science* 265:2088–2090.
- Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378:789–792.

APPENDIX

BIC Consortium Steering Committee

Current members:

Anne-Lise Boeressen, Norway
Lawrence C. Brody, USA
Fergus J. Couch, USA
Graham Casey, USA
Peter Devilee, The Netherlands
Douglas Easton, United Kingdom
Charis Eng, USA
Simon Gayther, United Kingdom
David Goldgar, France
Kate Nathanson, USA
Susan Neuhausen, USA
Elaine Ostrander, USA

Kelly Owens, USA
Csilla Szabo, France
Sean Tavtigian, USA

Past members:

Stephen Friend
Andrew Futreal
Patricia Murphy
Mark Rabin
Elizabeth Schubert
Donna Shattuck-Eidens
Barbara Weber
Roger Wiseman